JOURNAL OF CHEMOMETRICS, VOL. 3, 329–342 (1989)

THE PERFORMANCE OF LEAST SQUARES AND ROBUST REGRESSION IN THE CALIBRATION OF ANALYTICAL METHODS UNDER NON-NORMAL NOISE DISTRIBUTIONS

R. WOLTERS AND G. KATEMAN

Department of Analytical Chemistry, University of Nijmegen, Toernooiveld, NL-6525 ED Nijmegen, The Netherlands

SUMMARY

By means of Monte Carlo simulations a comparison has been made between ordinary least squares regression and robust regression. The robust regression procedure is based on the Huber estimate and is computed by means of the iteratively reweighted least squares algorithm. The performance of both procedures has been evaluated for estimation of the parameters of a calibration function and for determination of the concentration of unknown samples. The influence of the distributional characteristics skewness and kurtosis has been studied, and the number of measurements used for constructing the calibration curve has also been taken into account. Under certain conditions robust regression offers an advantage over least squares regression.

KEY WORDS Calibration Robust regression Iteratively reweighted least squares M-estimator Monte Carlo simulations

INTRODUCTION

In chemical analysis the analytical methods usually have to be calibrated before they can be used to estimate some property (often the concentration of a component) of unknown samples. For this purpose the response of the method is measured for a number of samples for which the considered property is known. On the basis of these measurements a model is calculated that relates the response to the concentration. Next, this relation can be used to estimate the concentration of an unknown sample after its response has been measured. So in fact two steps are performed: the modelling of the response and the prediction of the unknown concentrations.

Least squares regression is a very popular and widespread statistical procedure for obtaining the relation between response and concentration. The least squares estimate is utilized in many regression procedures. It is used for univariate as well as multivariate regression. The structural form of the model may be linear (e.g. straight line, polynomial) or non-linear in the parameters, in which cases the procedures are called linear or non-linear least squares regression respectively. For least squares regression it is assumed that the noise in the responses is identically distributed over the concentration range and that the noise terms are

0886–9383/89/020329–14$07.00

Received 21 March 1988
Revised 29 June 1988

independent of each other. When the noise is not identically distributed, a weighted least squares estimate is appropriate; and when the noise is also not independent, i.e. the responses are correlated, a generalized least squares estimate is available.[1]

An important assumption for all these procedures is that the noise has a normal (or Gaussian) distribution. This assumption is based on the idea that the observed noise is the result of a large number of small independent errors, resulting from e.g. manipulative operations and electrical noise. In analytical practice, however, these conditions may not be met, which can result in non-normal distributions.[2] Clancey[3] examined some 250 error distributions involving about 50 000 chemical analyses of metals. It was shown that only about 10–15% of the distributions could be regarded as normal. About 20% were symmetrical, but showed leptokurtosis (kurtosis >3), 20–25% were skew distributions and the remainder were of another shape or were irregular. Non-normal distributions were also found in analyses of various blood constituents.[4] The observed deviations of the normal distributions may be due to the inherent properties of the analytical method or may be caused by the presence of one or more outliers. The least squares estimate is very sensitive to outliers. An arbitrary gross error has an unbounded influence on the results. Therefore the least squares estimate is regarded as non-robust with respect to the distribution.

Robustness can be defined as the insensitivity to small deviations from the assumptions. In this case we are concerned with the distributional robustness, which means that the true underlying distribution deviates slightly from the normal distribution. A robust procedure should possess the following features:[5]

1. It should be reasonably efficient for the assumed model.
2. Small deviations from this model should impair the performance only slightly.
3. Larger deviations should not cause a catastrophe.

The problems associated with outliers and robustness were recognized early in the 19th century, but a systematic approach dates only from the last 20 years. More recently a number of textbooks on the subject have been published.[5–8] In most of these textbooks the emphasis is on a class of robust estimators that are called maximum likelihood estimators (M-estimators). This is a type of estimator that minimizes a different function of the residuals than the sum of squares, e.g. the sum of absolute residuals.

Applications in analytical chemistry of robust regression based on M-estimators have been few. Phillips and Eyring[9] applied an M-estimator, the biweight function of Tukey, in regression analysis of analytical data. They showed that in most cases the efficiency of the robust regression was about the same or better than the least squares regression. Massart *et al.*[10] applied the LMS-estimate (least median of squares) to a number of experimental data from the chemical literature. LMS-regression is a more recently developed robust method, described by Rousseeuw *et al.*,[11] in which minimizing the sum of squares of residuals is replaced by minimizing the median of squares of residuals. The method is very resistant for a high outlier rate. Up to 50% outliers can be tolerated, because only the best half of the measurements is used for estimating the regression coefficients.

In both papers[10,11] experimental data from the literature have been used, for which the noise distributions are unknown. In this paper the results of Monte Carlo simulations will be presented. The influence of the shape of the distribution has been studied by using symmetrical distributions with long tails as well as skew distributions. Also the influence of the number of measurements used for calibration has been investigated. In this study the M-estimator as proposed by Huber[12] was used. This relatively simple estimator resembles the classical sum of least squares estimator. It differs in the fact that lower weights are assigned to

measurements that deviate more from the fitted line. This concept renders the estimator robust for outliers in the dependent variable (the response). M-estimators are not robust with respect to outliers in the independent variable (the concentration), so-called leverage points. For calibration purposes this is not a necessity, because the design of the concentration levels is controlled by the experimenter. Calibration designs should not contain leverage points; usually a design of equally spaced concentration levels is applied. However, in situations where outliers in the independent variable are expected, a method should be used that is also robust in that aspect, such as GM-estimators or the LMS-estimator.

Huber's M-estimator has been studied extensively in the statistical literature, under symmetric[13] as well as asymmetric[14] distributions. These Monte Carlo studies were carried out with sample sizes of 20 or more. This is a rather impractical number of calibration measurements for most analytical purposes, therefore sample sizes ranging from 3 to 20 will be investigated here.

THEORY

In this section the theory of M-estimators will be outlined. Because the ordinary least squares estimate is in fact also a member of this class of estimates, it will be treated as well, albeit in a slightly different way than usual.

The objective of regression is to fit the experimental data to a model. Here a linear model will be assumed:

$$y_i = \sum_{j=1}^{p} (x_{i,j} * \beta_j) \tag{1}$$

with y_i the ith measured response, $x_{i,j}$ the (i,j)th element of the design matrix and β_j the jth parameter of the model. The variance in the measured response is assumed to be constant and independent of x. The fit of the data on the model is obtained by minimizing the sum of some function (ρ) of the residuals (r):

$$r_i = y_i - \sum_{j=1}^{p} (x_{i,j} * \beta_j) \tag{2}$$

$$\sum_{i=1}^{n} \rho(r_i) = \min! \tag{3}$$

or, after taking the derivatives with respect to all β,

$$\sum_{i=1}^{n} \phi(r_i) * x_{i,k} = 0 \quad \text{for } k = 1, \ldots, p \tag{4}$$

In order to make expression (4) invariant to scale, an auxiliary estimate of scale (S) is introduced:

$$\sum_{i=1}^{n} \phi(r_i/S) * x_{i,k} = 0 \quad \text{for } k = 1, \ldots, p \tag{5}$$

So far, the theory is valid for robust regression with M-estimates as well as for least squares regression. The difference between the two procedures lies in the choice of the ρ- and ϕ-functions. For least squares regression these are defined as

$$\rho = \tfrac{1}{2} r^2 \tag{6}$$

$$\phi = r \tag{7}$$

With these functions expression (5) turns into the following set of equations (the scale estimate can be omitted here, because in this case the expression is scale-invariant):

$$\sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} (x_{i,j} * \beta_j) \right) * x_{i,k} = 0 \quad \text{for } k = 1, \ldots, p \tag{8}$$

This is a set of p equations with p variables and its solution is straightforward. Written in matrix notation this leads to the well known solution for β in the case of linear least squares regression:

$$\beta = (\mathbf{X}^{\mathrm{T}} \mathbf{X})^{-1} \mathbf{X}^{\mathrm{T}} \mathbf{Y} \tag{9}$$

For robust regression with the Huber estimate the ρ- and ϕ-functions are more complicated:[15]

$$\rho = \begin{cases} \frac{1}{2} r^2 \\ k*|r| - \frac{1}{2} k^2 \end{cases} \quad \text{for} \quad \begin{cases} |r| \leqslant k*S \\ |r| > k*S \end{cases} \tag{10}$$

$$\phi = \begin{cases} r \\ k*|r|/r \end{cases} \quad \text{for} \quad \begin{cases} |r| \leqslant k*S \\ |r| > k*S \end{cases} \tag{11}$$

For small residues the functions are the same as for the least squares estimate, but for larger values of r the ρ-function is modified in such a way that the contribution to the minimization criterion is reduced. In this case the scaling factor can no longer be omitted and a suitable robust estimate for scale has to be used. Usually the median of absolute deviations (MAD) is used. The constant k in (10) and (11) is called the tuning constant. The value of k determines at which point the weight of the residues is reduced and hence determines the robustness of the estimate. At very large k-values the estimate is equivalent to the least squares estimate. A very small k-value gives an estimate that is very robust but has a low efficiency. Here a value is chosen of $k = 1{\cdot}345$. For this k-value the asymptotic relative efficiency is 95%, which means that under normal distributions the variance of the least squares estimate is 95% of the variance of the robust estimate.[16]

The solution of equation (5) is no longer straightforward. The way used most often to solve the problem is by means of an iteratively reweighted least squares (IRLS) procedure. The solution vector of β is defined as

$$\beta = (\mathbf{X}^{\mathrm{T}} \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^{\mathrm{T}} \mathbf{W} \mathbf{Y} \tag{12}$$

where $\mathbf{W}$ is an n-dimensional vector of weights. The weights are a function of the residulas and of the scaling factor. Generally they are defined as $w_i = \phi(r_i)/r_i$. For the Huber estimate this results in

$$w_i = \begin{cases} 1 \\ k/|r_i| \end{cases} \quad \text{for} \quad \begin{cases} |r_i| \leqslant k*\mathrm{MAD} \\ |r_i| > k*\mathrm{MAD} \end{cases} \tag{13}$$

This function is shown in Figure 1. The iteration procedure consists of consecutive estimation of the MAD and, after recalculation of the weights, estimation of β. Iteration continues until convergence is reached.

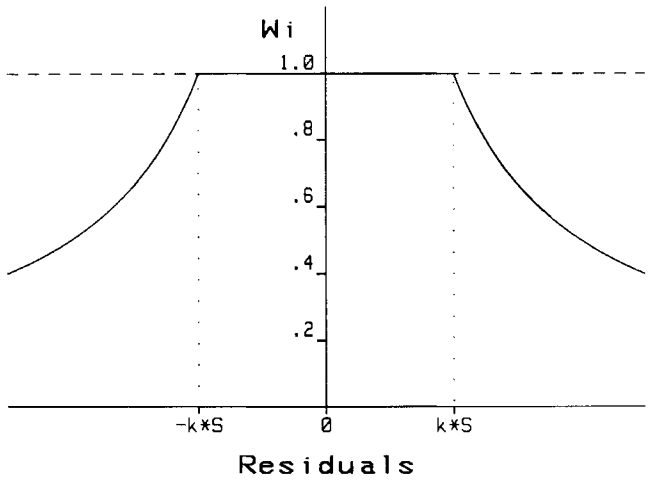

Figure 1. The weight function used in the iteratively reweighted least squares procedure for the Huber estimate (——) and the least squares estimate (– – –)

EXPERIMENT

The calculation of the Huber estimate involves an iterative procedure. Hence a set of starting values is needed. The least squares estimate is a possible option, but because of its non-robustness this might lead to incorrect results. Therefore a correction has to be made to the least squares estimate.[9] The residuals are Winsorized according to

$$r_i^w = \begin{cases} -1{\cdot}5*\text{MAD} \\ r_i \\ 1{\cdot}5*\text{MAD} \end{cases} \quad \text{for} \quad \begin{cases} r_i < -1{\cdot}5*\text{MAD} \\ |r_i| \leqslant 1{\cdot}5*\text{MAD} \\ r_i > 1{\cdot}5*\text{MAD} \end{cases} \tag{14}$$

The vector $\mathbf{R}^w$ is used in expression (9) instead of $\mathbf{Y}$ to obtain a p-dimensional correction vector which is added to the least squares estimate. This correction reduces the effects of outliers and provides satisfactory starting values. The iterations were stopped when the relative change in the estimated parameters was less than $1{\cdot}0 \times 10^{-6}$.

Table 1. Parameters of the generated noise distributions

Distr. no.	R	δ	c	Variance	Skewness	Kurtosis
1	0	0	1	1·00	0	3·00
2	0·01	0	3	1·08	0	4·63
3	0·05	0	3	1·40	0	7·65
4	0·1	0	3	1·80	0	8·33
5	0·2	0	3	2·60	0	7·54
6	0·01	2	1	1·04	0·073	3·14
7	0·05	2	1	1·19	0·263	3·38
8	0·1	2	1	1·36	0·363	3·36
9	0·2	2	1	1·64	0·366	3·04

For the Monte Carlo experiments nine different noise distributions were generated, all of which satisfy the equation of a contaminated normal distribution:

$$\text{noise} = (1-R)*N(0, \sigma^2) + R*N(\delta, c^2\sigma^2) \tag{15}$$

The coefficients for these distributions are tabulated in Table 1, together with their asymptotic variance, skewness and kurtosis. Figure 2 shows the shape of these distributions. The first distribution has a fraction contamination of zero and in fact is the normal distribution. Distributions 2–5 are of symmetrical shape with more heavy tails, and distributions 6–9 are skew distributions.

The simulation experiments on calibration were performed with each of these distributions. For the calibration curve the following model was chosen:

$$y = \beta_1 + \beta_2 * x \tag{16}$$

with y the analytical response, x the analyte concentration, β_1 the intercept of the calibration-line and β_2 the slope of the calibration-line.

Because the values of the parameters β_1 and β_2 are not critical in this situation, they were both set to 1·0. The measurement of a calibration standard was generated by calculating its response from equation (16) and superimposing the noise of the desired distribution with a variance of $\sigma^2 = 0{\cdot}001$. The different numbers of calibration standards were 3, 4, 5, 6, 8, 10, 15 and 20. In every case the standards were evenly distributed over the concentration range, which was set from 0·0 to 1·0. With such a series of simulated measurements a calibration was performed, using one of the regression procedures described above. For each calibration line a series of five measurements, at concentrations of 0·00, 0·25, 0·50, 0·75 and 1·00, were also simulated. For these measurements the corresponding concentrations were determined using the estimated calibration line. The predicted concentrations, as well as the parameters of the calibration line, were stored for each number of observations. The simulations were repeated 2000 times per number of observations and per distribution. The same random numbers were used for every distribution in order to reduce the variation in the results.

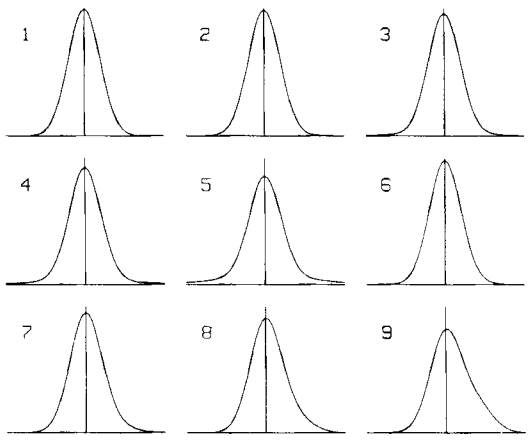

Figure 2. Plot of the generated noise distributions

All simulations and calculations were performed on an IBM-AT personal computer equipped with an 80287 mathematical coprocessor. The necessary programs were written in Turbo Pascal (Borland).

RESULTS AND DISCUSSION

In the results of the simulations a distinction can be made between the modelling performance and the reliability of the predicted concentrations.

First the results concerning the modelling performance will be presented. A suitable measure for this performance is the relative efficiency with which the parameters of the calibration line can be estimated. The relative efficiency is defined as the quotient of the variance under the considered conditions and the variance under reference conditions:

$$E_{ref} = \sigma^2_{ref}/\sigma^2 \tag{17}$$

with E_{ref} the relative efficiency, σ^2_{ref} the variance under reference conditions and σ^2 the variance under the considered conditions.

In the case of ordinary least squares regression the efficiency of the procedure is determined

Table 2. Efficiency of parameter estimation by robust regression relative to least squares regression

Parameter 1 (intercept)

Distr. no.	Number of measurements 3	4	5	6	8	10	15	20
1	0·826	0·900	0·874	0·858	0·888	0·899	0·881	0·878
2	0·828	0·921	0·899	0·890	0·932	0·954	0·919	0·921
3	0·834	0·940	0·971	1·026	1·058	1·084	1·105	1·098
4	0·856	0·965	1·002	1·110	1·197	1·235	1·285	1·264
5	0·851	0·961	1·047	1·150	1·261	1·319	1·452	1·460
6	0·823	0·906	0·883	0·869	0·902	0·911	0·881	0·882
7	0·814	0·914	0·891	0·882	0·913	0·930	0·911	0·909
8	0·819	0·934	0·885	0·900	0·930	0·947	0·915	0·935
9	0·821	0·920	0·869	0·861	0·886	0·889	0·889	0·903

Parameter 2 (slope)

Distr. no.	Number of measurements 3	4	5	6	8	10	15	20
1	1·000	0·952	0·870	0·866	0·881	0·894	0·892	0·873
2	1·000	0·956	0·896	0·894	0·907	0·943	0·947	0·915
3	1·000	0·958	0·956	1·009	1·040	1·061	1·111	1·079
4	1·000	0·957	0·981	1·078	1·166	1·183	1·277	1·254
5	1·000	0·949	1·019	1·110	1·257	1·299	1·454	1·441
6	1·000	0·953	0·872	0·878	0·896	0·915	0·897	0·876
7	1·000	0·951	0·878	0·893	0·908	0·938	0·924	0·910
8	1·000	0·952	0·876	0·922	0·920	0·948	0·922	0·935
9	1·000	0·951	0·868	0·878	0·899	0·900	0·910	0·910

for all distributions relative to the normal distribution. In Figure 3 this relative efficiency (E_N) is depicted as a function of the fraction contamination (R) for all the different numbers of observations. The data for the symmetrical and skew distributions are plotted separately (Figure 3(a) and Figure 3(b) respectively). The efficiency of the variance of the parameters decreases approximately proportionally to the reciprocal of the asymptotic variance, as listed in Table 1. This is in accordance with what could be expected.

In the case of robust regression with the Huber estimate the efficiency is determined relative to the results with the same distribution in the case of least squares regression. In Table 2 this efficiency (E_{LSE}) is shown for both parameters. In Figure 4 the results are plotted as a function of the number of observations. Again the data for the symmetrical and skew distributions are plotted separately (Figure 4(a) and Figure 4(b) respectively).

Figure 4(a) shows some interesting features of the robust regression procedure. First the behaviour for small numbers of observations will be discussed. When only two measurements are made to estimate the calibration line, no residues exist and consequently a robust regression algorithm is not feasible. For three measurements per calibration line it is possible to perform a robust regression procedure. However, during the iterations the weight of the middle point converges to zero and the weights of the two outer points become one. This results in a value for the slope that is the same as that found by least squares regression. Therefore the relative efficiency for the second parameter is 1·0. The estimate for the intercept will differ from that found by least squares regression. It will be larger or smaller, depending on whether the middle point has a negative or positive residue in least squares regression. This feature results in a lower value for the efficiency of the first parameter. The effect is reduced when the number of measurements is increased. For five or more measurements the efficiencies for both parameters become approximately the same. From that point onwards a distinction becomes clear between the different distributions (different lines). The line for the normal distribution is below 1·0 over the plotted range. This was to be expected because the least squares estimate is the most efficient one in the case of a true normal distribution. But when the distribution deviates more and more from the normal distribution (R increases), the relative efficiency of robust regression increases and grows beyond 1·0. When the efficiency is over 1·0, robust regression is more efficient than the least squares estimate and therefore results in better estimates for the calibration line. The point of transition is dependent on the fraction contamination of the normal distribution and on the number of measurements. The points where the transition occurs can be estimated from Figure 4(a) for both parameters. The interpolated points are depicted in Figure 5. The area above the curve in this figure represents the combination of the fraction contamination and the number of measurements where the robust Huber estimate results in better values for the parameters.

The results for the skew distributions (Figure 4(b)) are similar to those for the symmetrical distributions for small numbers of measurements (less than five). For larger numbers of measurements all lines for the different skew distributions almost coincide. Only small enhancements of the modelling performance can be detected. This is most probably caused by the fact that the Huber estimate was designed for symmetrical distributions.

After discussing the modelling performance, we will take a look at the reliability of the predicted concentrations. This reliability comprises two components: the systematic error (bias) and the random error (precision). Both error components are combined in the criterion of the total error, which is defined as[16]

$$T = \text{abs}(\delta) + 2*\sigma \tag{18}$$

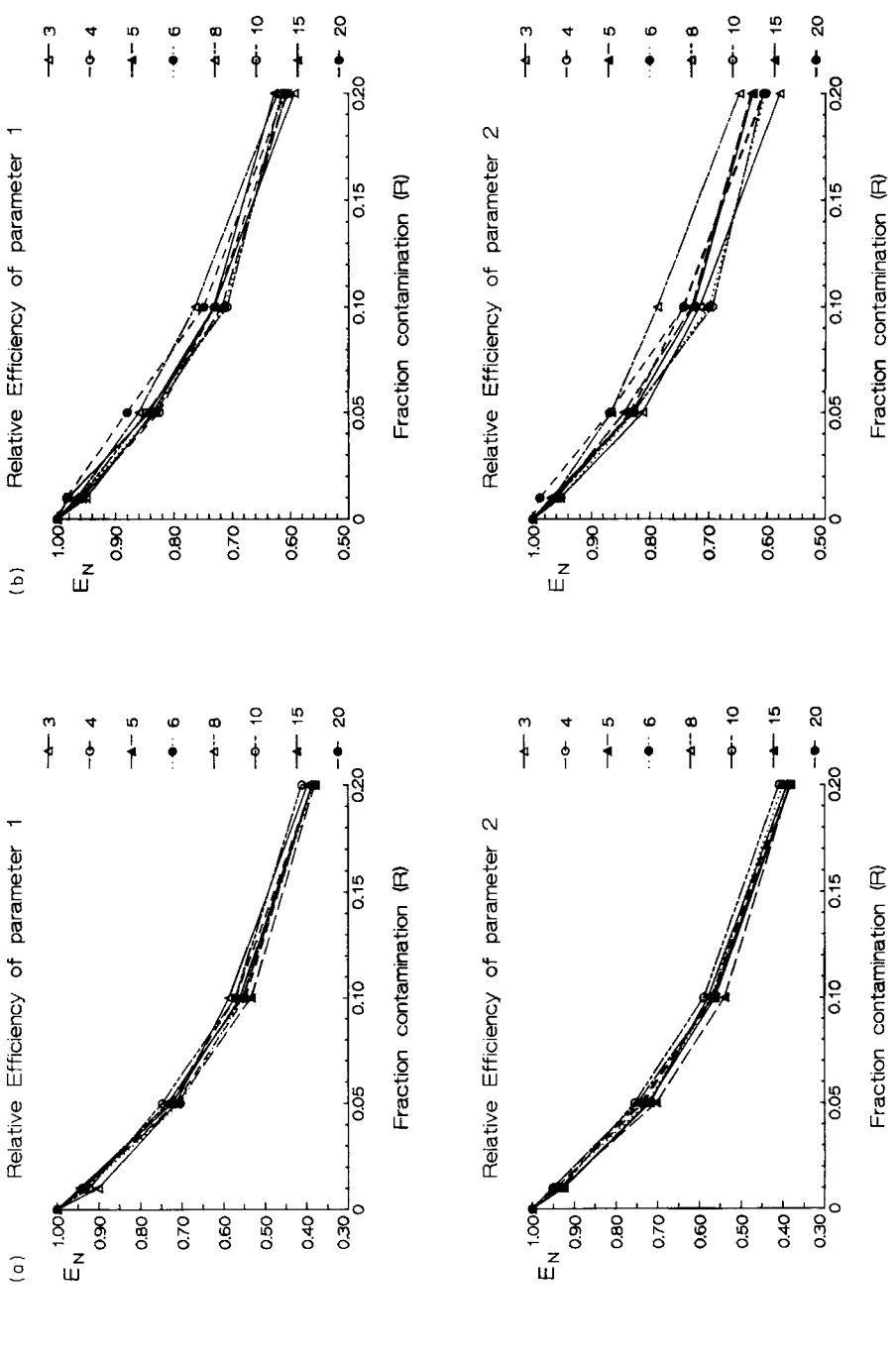

Figure 3. The efficiency of the least squares estimate relative to the normal distribution (E_N) as a function of the fraction contamination (R) for all eight numbers of measurements: (a) for the symmetrical distributions (both parameters); (b) for the skew distributions (both parameters)

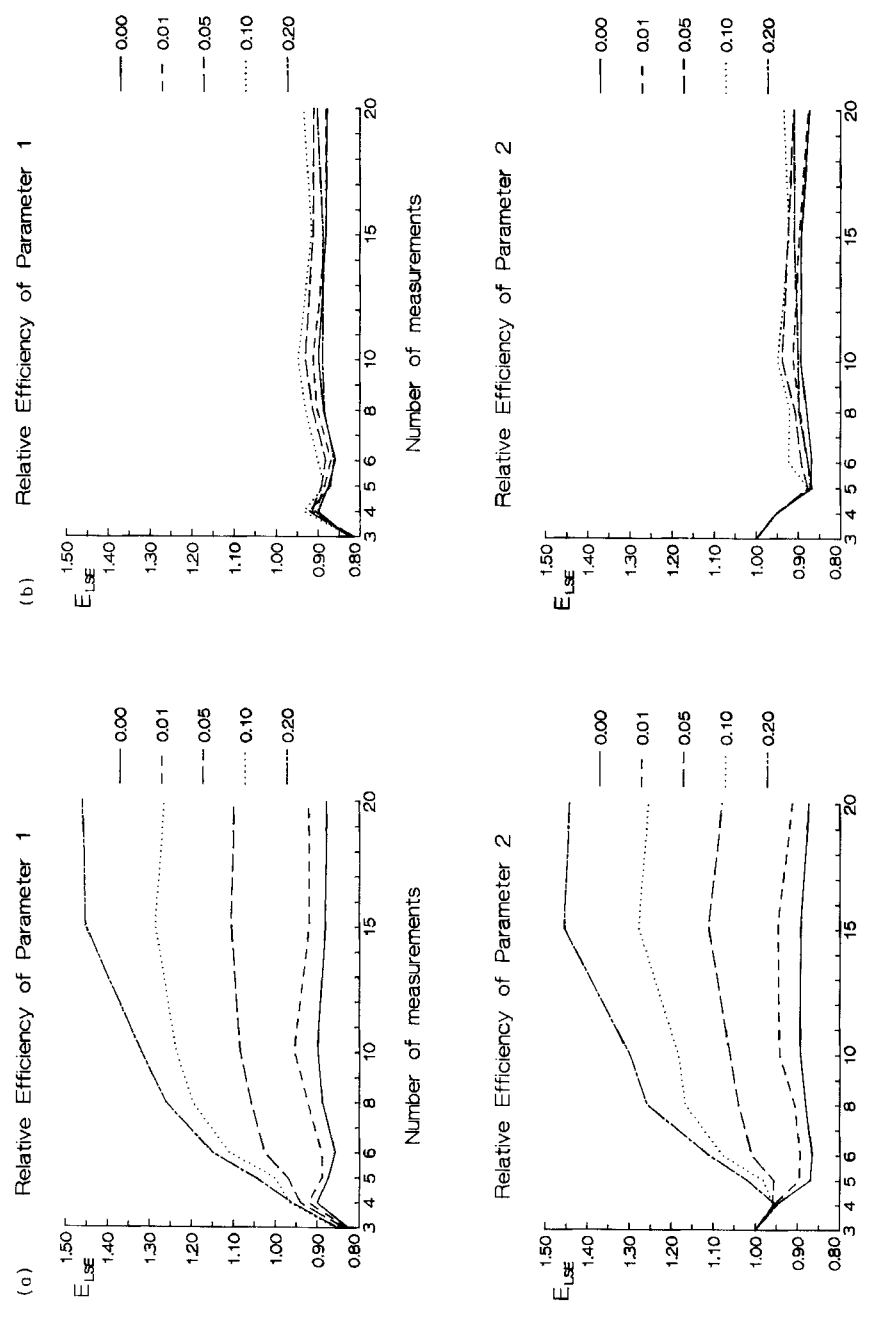

Figure 4. The efficiency of robust regression relative to least squares regression on the same noise distribution (E_{LSE}) as a function of the number of measurements for the five different values of the fraction contamination: (a) for the symmetrical distributions (both parameters); (b) for the skew distributions (both parameters)

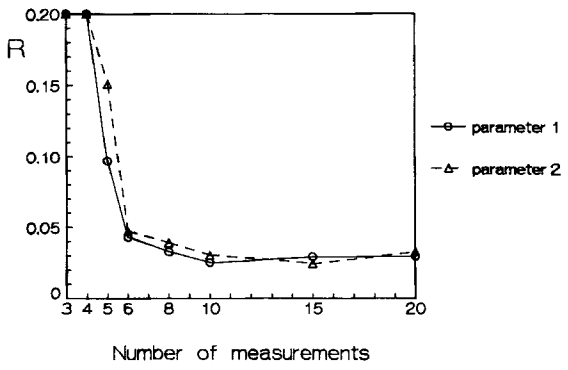

Figure 5. The line of equal efficiency of robust regression and least squares regression for both parameters. In the region above the curve robust regression is more efficient

with δ the systematic error (mean deviation of the true value) and σ the random error (standard deviation of δ). The factor 2 represents a confidence interval of approximately 95%. For this situation the total error is a suitable criterion, because all results are based on the same number of simulations and therefore the errors in δ and σ are the same for all results. If the results had been based on different numbers of simulations or, as in real life, on different numbers of repeated measurements, then a more complicated criterion would be used, such as the maximum total error.[17] The total error is calculated for each of the five predicted points. This is done for robust regression as well as for the least squares regression. The quotient of these two numbers gives the total error of robust regression relative to least squares regression. Finally, quotients for the five predicted points are averaged, which results in the average relative total error (ART)

$$\text{ART} = \tfrac{1}{5} * \sum_{i=1}^{5} (T_{\text{rob},i}/T_{\text{lse},i}) \tag{19}$$

with $T_{\text{rob},i}$ the total error for robust regression for point i and $T_{\text{lse},i}$ the total error for least squares regression for point i. This criterion is calculated for all distributions and for all numbers of measurements. The results are shown in Table 3 and Figure 6. The behaviour of

Table 3. Average efficiency of determined concentrations by robust regression relative to least squares regression

Distr. no.	Number of measurements 3	4	5	6	8	10	15	20
1	1·051	1·020	1·020	1·022	1·015	1·009	1·007	1·010
2	1·052	1·018	1·015	1·018	1·009	1·003	1·000	1·007
3	1·053	1·008	1·005	0·999	0·995	0·994	0·990	0·996
4	1·046	1·006	0·998	0·991	0·984	0·985	0·981	0·992
5	1·047	1·004	0·992	0·986	0·976	0·980	0·975	0·991
6	1·050	1·019	1·020	1·021	1·015	1·009	1·006	1·009
7	1·054	1·018	1·023	1·020	1·018	1·013	1·009	1·016
8	1·050	1·019	1·028	1·021	1·020	1·018	1·018	1·022
9	1·055	1·021	1·034	1·029	1·028	1·027	1·026	1·027

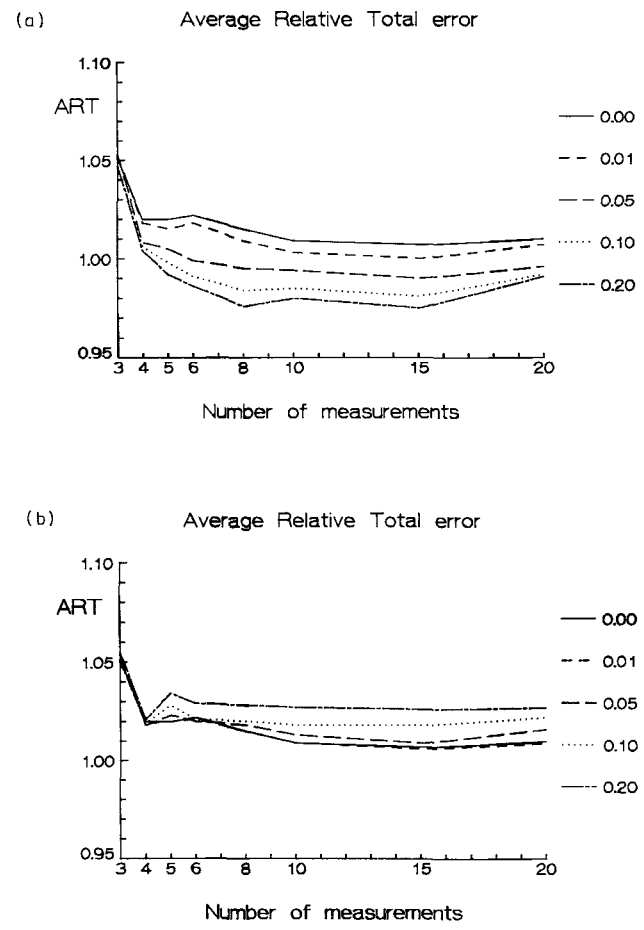

Figure 6. The average relative total error (ART) as a function of the number of measurements for the five different values of the fraction contamination: (a) for the symmetrical distributions; (b) for the skew distributions

the predicted concentrations is similar to that of the modelling efficiency. For a small number of measurements per calibration line the errors in the prediction of unknown samples are always larger when using robust regression than when using least squares regression. For the symmetrical distributions robust regression gives smaller total errors than least squares regression for a number of measurements of six or more. However, the advantage of the robust procedure has become less. This is caused by two factors. First, there is a difference between the two criteria (18) and (19). The main contribution to the ART is the standard deviation (σ). In the efficiency the variance is used, which is the square of the standard deviation. This causes the differences in efficiency to be more pronounced. Second, although the effect of outliers is reduced by robust regression, the outliers are still present in measurements of unknown samples. This causes an increase in the standard deviation of the predicted concentrations and partially masks the better modelling performance of robust regression. For skew distributions there is still almost no distinction between the different distributions, and least squares regression is superior.

CONCLUSIONS

In chemical analysis the measurement noise cannot always be considered to be normal. Some noise distributions are symmetrical but have heavier tails than the normal distribution. For this kind of noise it can be advantageous to use a robust regression procedure based on an M-estimate. For the Huber estimate it is shown here that, for a number of measurements larger than ten, the use of the robust procedure results in a better estimate of the calibration model when the fraction contamination (R) is over 0·035. This value of R represents an asymptotic kurtosis of approximately 6·6. However, for a small number of calibration measurements (less than five) the least squares estimate always gives a more precise estimate. Many actual noise distributions were found to be asymmetrical. For these distributions the use of a robust procedure will not help. In that case a commonly suggested solution is a transformation of the data to a symmetrical distribution, for instance by means of a power transformation.[18] It should be borne in mind though that a transformation to symmetry also changes originally homoscedastical data to heteroscedastical. Consequently, weighted least squares regression is appropriate for the transformed data.

Although the estimation of the calibration line can be greatly improved by robust regression in the case of heavily tailed distributions, the outliers present in the measurements of unknowns affect the total errors. This means that a careful examination of the results is still necessary to prevent occasional gross errors in the results.

It should be noted that the above results are only shown to be valid for the used noise distributions and the Huber estimate with the used tuning constant. However, simulations with the biweight estimate showed similar results (not reported here). Also, the application of robust regression on actual chemical data sets showed approximately equal performance to or better performance than least squares regression.[9]

For true normal distributions ordinary least squares regression is no doubt the best choice. Using the robust procedure in this case will increase the total error of the predicted concentrations only by about 2%. Unfortunately in real life the shape of the distribution is rarely known. In some cases historical data may be available which can be utilized to estimate skewness and kurtosis of the distribution. Most of the time, however, there is no certainty about the distributional behaviour of the chemical measurement. In that case, for calibration designs that consist of five or less measurements, the use of the robust procedure is not recommended, because this will deteriorate the total error of the results. However, for larger calibration designs, especially for ten or more measurements, it is worthwhile to consider the use of a robust regression procedure. A relatively small price is paid when the distribution is actually normal, but good protection against the influence of outliers is obtained when the distribution is not normal.

REFERENCES

1. N. Draper and H. Smith, *Applied Regression Analysis*, 2nd Edn, Wiley, New York (1981).
2. M. Thompson and R. J. Howarth, *Analyst* **101**, 690 (1976).
3. V. J. Clancey, *Nature* **159**, 339 (1947).
4. E. K. Harris and D. L. DeMets, *Clin. Chem.* **18**, 605 (1972).
5. P. J. Huber, *Robust Statistics*, Wiley, New York (1981).
6. D. C. Hoaglin, F. Mosteller and J. W. Tukey (eds), *Understanding Robust and Exploratory Data Analysis*, Wiley, New York (1983).
7. W. J. J. Rey, *Introduction to Robust and Quasi-Robust Statistical Methods*, Springer-Verlag, Berlin (1983).

8. F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, *Robust Statistics: The Approach based on Influence Functions*, Wiley, New York (1986).
9. G. R. Phillips and E. M. Eyring, *Anal. Chem.* **55**, 1134 (1983).
10. D. L. Massart, L. Kaufman, P. J. Rousseeuw and A. Leroy, *Anal. Chim. Acta* **187**, 171 (1986).
11. P. J. Rousseeuw, *J. Amer. Statist. Assoc.* **79**, 871 (1984).
12. P. J. Huber, *Ann. Math. Statist.* **35**, 73 (1964).
13. P. W. Holland and R. E. Welsch, *Commun. Statist.–Theor. Methods A* **6**(9), 813 (1977).
14. R. J. Carroll, *J. Amer. Statist. Assoc.* **74**, 674 (1979).
15. P. J. Huber, *Ann. Math. Statist.* **43**, 1041 (1972).
16. E. F. McFarren, R. J. Lishka and J. H. Parker, *Anal. Chem.* **42**, 358 (1970).
17. R. Wolters and G. Kateman, *Anal. Chim. Acta* **207**, 111 (1988).
18. G. E. P. Box and D. R. Cox, *J. R. Statist. Soc. B* **26**, 211 (1964).